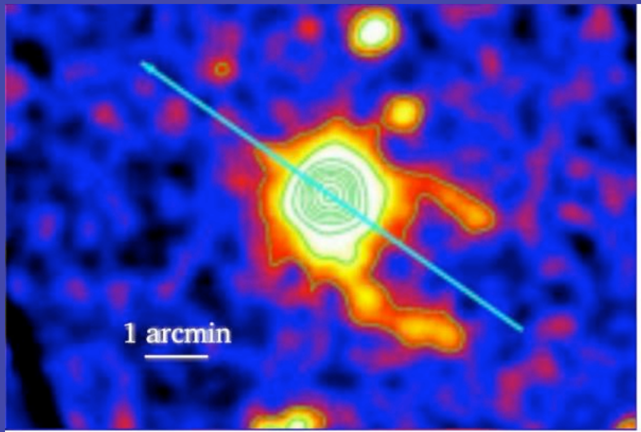
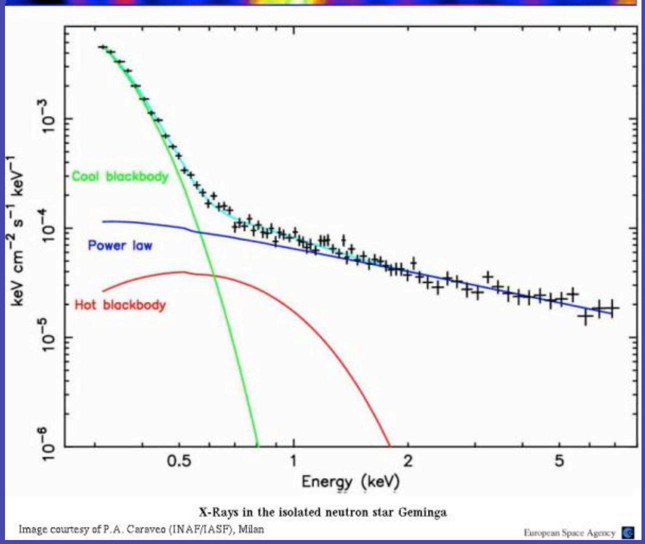
In the News...

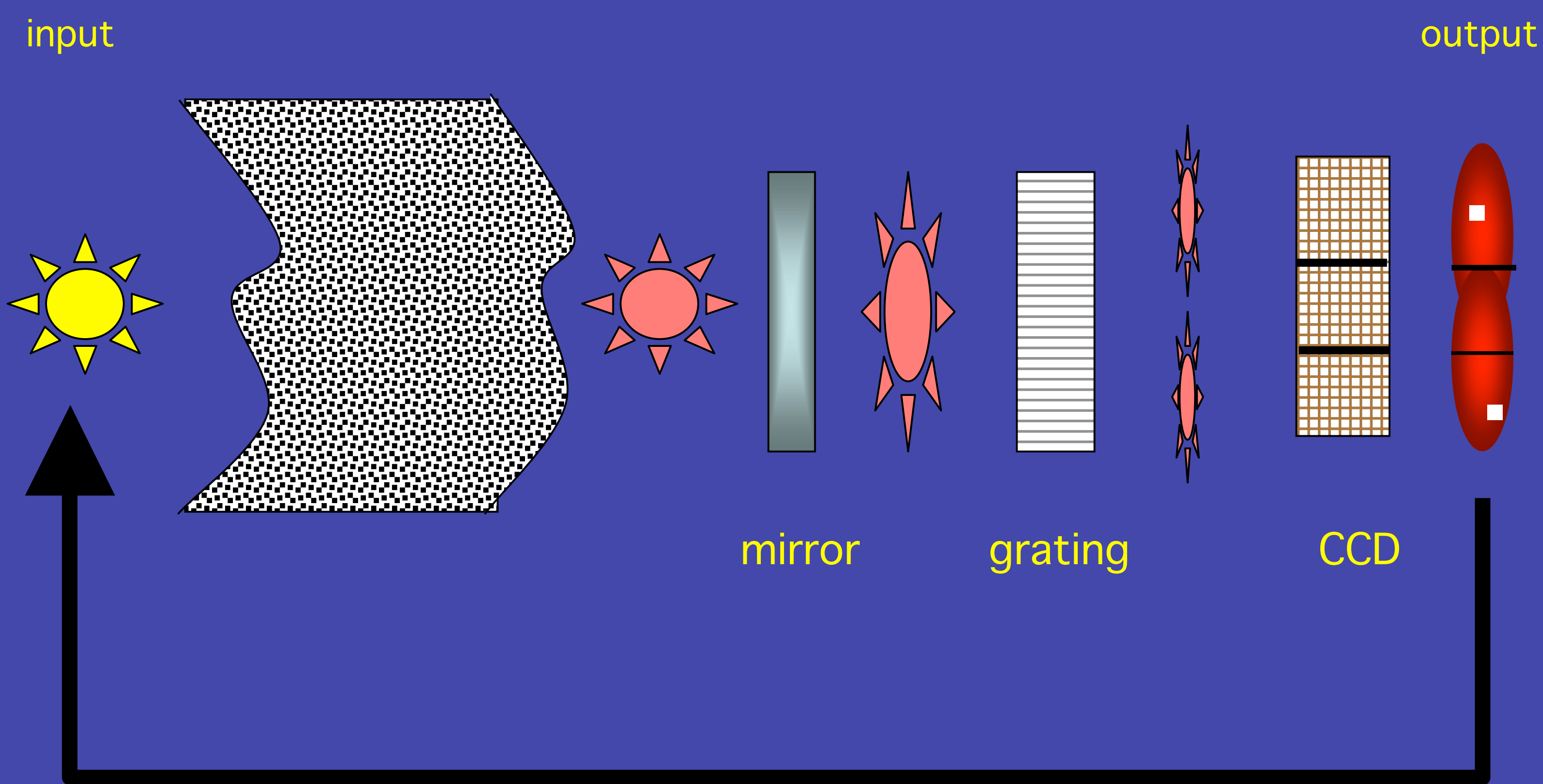




Upper: XMM-Newton image of Geminga (the nearest isolated neutron star at 500 pc), showing an X-ray "wake" due to its relatively large proper motion

Lower: XMM-Newton spectrum of Geminga, showing 3 individual components: A cool blackbody, a hotter blackbody, and a hard nonthermal component

Understanding Your Data



how to go from output to input...

Counting

High energy astronomy usually starts with the detection of an event. ("Photon Counting")

The essential question is *how does the observed count rate (events per second) recorded by the detector correspond to the source rate of photon emission*?

The Counting Relation

The fundamental counting relation between detected counts in some detector unit and the source flux is

$$C(I) = \int_{t_2}^{t_1} \int_0^{\infty} f(E)\, dt\, R(I,E)\, dE + B(I)$$

where

- C(I) is the observed number of counts in the i^{th} detector channel;
- B(I) is the detected number of background counts;
- f(E) is the flux incident on the instrument; and
- R(I,E) is the response of the instrument

All the source physics is contained in f(E); the instrumentation physics is contained in R(I,E) and B(I).

Types of Instrumentation

Recall that astronomical instrumentation consists of:

- light gathering device (mirror system, collimator)
- disperser (optional)
- filter (optional)
- photo-sensitive detector
    - ✓ Proportional counter
    - ✓ CCD
    - ✓ scintillators...

Effects of these on incident radiation must be understood & characterized

Imperfections must be identified and calibrated.

Effects of Instrumentation

Recorded data is affected by:

- ability to count fast enough
- ability to collect charge completely
- ability to determine the energy of the output pulse
- ability to reconstruct position of event on detector

Data Reduction

Data reduction is the process by which information from a detector is cleaned and corrected for instrumental imperfections

It involves the correction of observed events ("counts") for detector artifacts, non-linearities, and other imperfections.

$$\boxed{C(I)} = \int_{t_2}^{t_1} \int_0^{\infty} f(E)\, dt\, R(I, E)\, dE + B(I)$$

i.e. convert $C(I)_{observed}$ to C(I) for further analysis

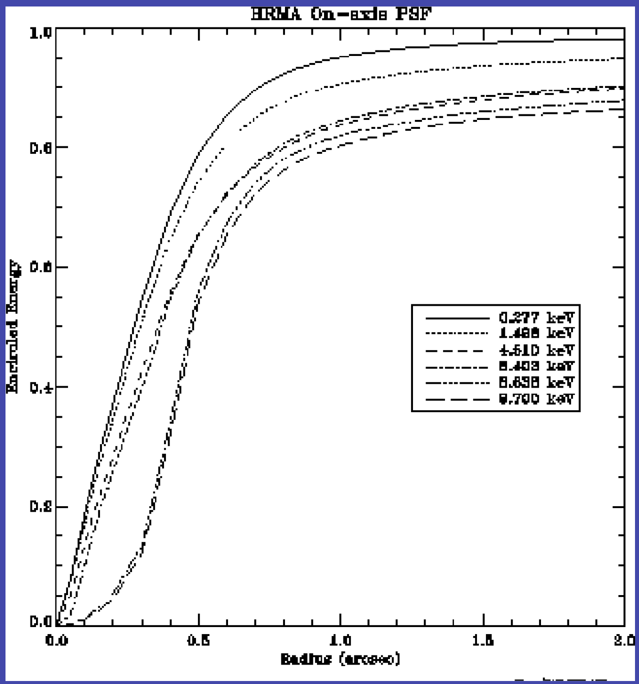
Reduction Issues

Instrumentation is not perfect.

1. Light gathering systems alter properties of incident radiation
2. Detector sensitivity varies with operating conditions and with detector properties (position on the detector, temperature, voltage levels, number of events being recorded…)
3. Detectors don’t respond instantaneously

“Correction” of Point-Source Spreading

Mirrors + detectors distribute photons from point sources over a finite area of the detector (i.e. over a finite number of pixels). The amount of spreading is characterized by the “point spread function”



PSF for the Chandra mirror, on-axis. Generally the PSF will increase with increasing off-axis angle

Gain Correction

Energy-sensitive detectors usually record the “pulse height” of an event. The pulse height is proportional to the photo-electron charged induced by event, which is proportional to the energy of the photon. A detector assigns a quantized pulse height to each event.

The detector gain is the relation between detector pulse height and input photon energy.

In general a detector’s gain is non-uniform and varies over the area of the detector. Before events from one region of the detector can be compared to events from another region, the variation must be corrected. One such correction involves use of a “gain map” which describes how the gain varies vs. position on the detector.

Correction for Deadtime

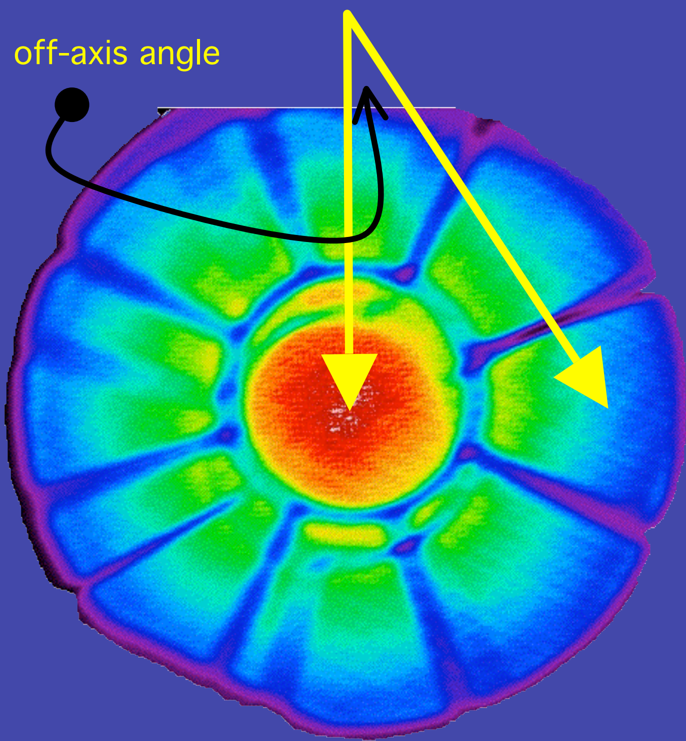
Detectors don't respond instantaneously to stimuli. There's some delay between the arrival time of a photon and the response of the detector.

Deadtime: There is usually some time when a detector processes an event during which time the detector is insensitive to the arrival of another event. If an observation spans a time Δt, and the Detector deadtime is δt, and N events are detected, then the corrected event rate seen by the detector is

$$R = N/(\Delta t - \delta t)$$

Vignetting Correction

Vignetting refers to the fact that mirror+ detector responses vary with off-axis angle, so that as you move away from the center of the mirror (away from the optical axis) the mirror becomes less efficient at focusing radiation.



Vignetting can be quantified using an "exposure map", a map of the relative amount of exposure (seconds in each pixel) during a given observation.

red: High exposure
blue: low exposure

Barycentric Correction

The photon arrival time is in a local frame (either the frame of the earth or the frame of the satellite). The observing frame is moving, so

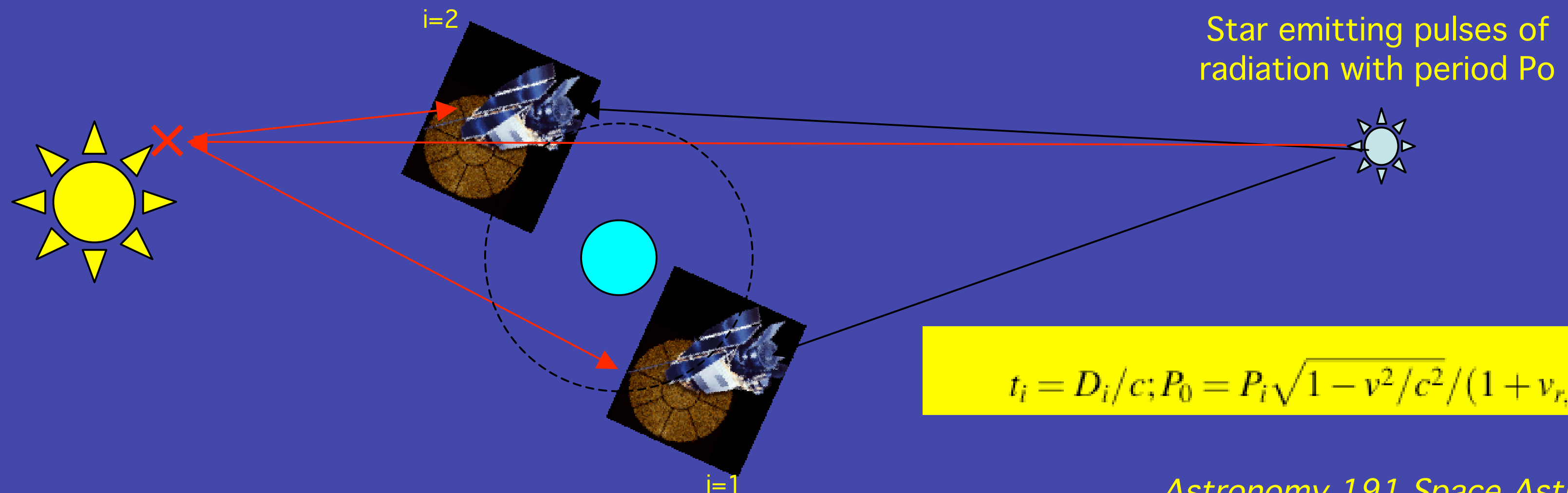
a) the distance between the source and the satellite is changing and

b) the satellite is moving towards or away from the source. This causes different absolute arrival times of photons at the location of the satellite via the doppler shift.

Generally this effect is small but must be taken into account if you are interested in very short time periods (like the pulse times of neutron star, ~milliseconds)

Barycentric correction: the arrival time of a photon is shifted as is it would have been detected at the barycenter (center of mass, ×) of the solar system instead at the position of the satellite.

see http://wave.xray.mpe.mpg.de/xmm/cookbook/EPIC_MOS/timing/barycen.html



$$t_i = D_i/c; P_0 = P_i\sqrt{1-v^2/c^2}/(1+v_{r,i})$$

CCD Problems

CCD's have the following imperfections:

- Dead Pixels - no sensitivity
- Hot Pixels - too much sensitivity
- Saturation & Bleeding
- Dark Current
- Pileup

Effects of Detector Pileup

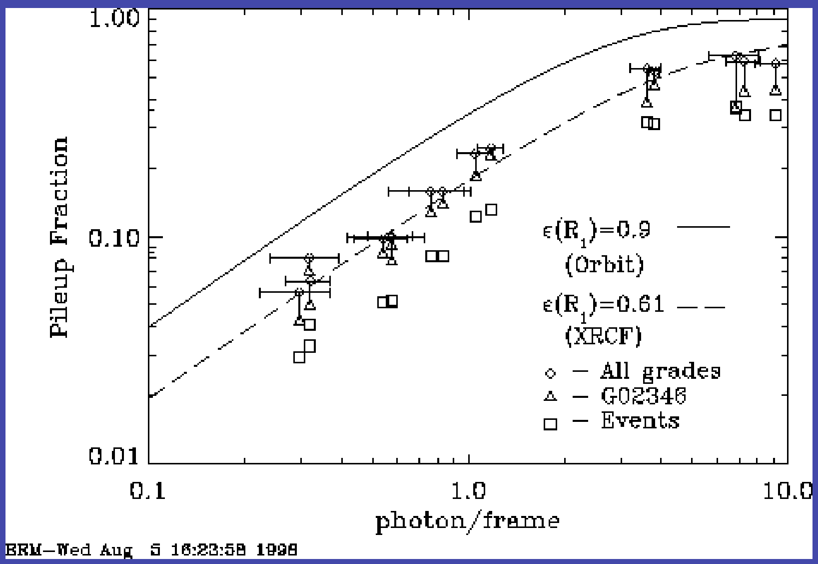
Pileup: in CCD detectors, events in a given pixel “wait” in a special region of memory (the “frame store”) and are read out. If another event strikes that pixel, that induced charge gets added to the frame store. Result:

— 2 events get counted as a single event (reduces count rate)

pileup fraction = (Number of piled-up events)/(Total number of detected events)

— recorded energy of combined event is

$$E_{combined}=E_1+E_2$$



pileup fraction vs. counting rate for ACIS

Data Analysis

Data analysis involves the comparison of some physical model to the corrected counts.

$$C(I) = \boxed{\int_{t_2}^{t_1} \int_0^{\infty} f(E)\,dt\,R(I,E)\,dE} + B(I)$$

The observed counts are a convolution of the source emission (the physical model) with the response of the instrument.

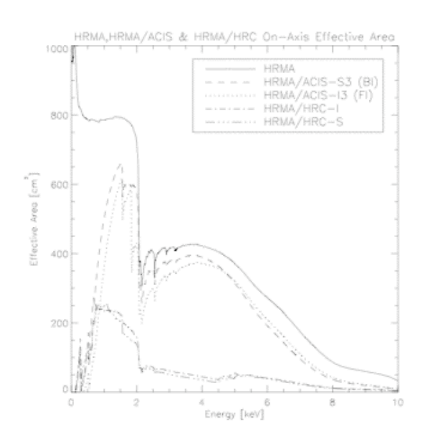
- Sometimes the instrument response is simple and can be deconvolved from the observed counts for a direct comparison to the physical model.

- High energy detectors have complicated responses, so deconvolution is usually not practicable or possible.

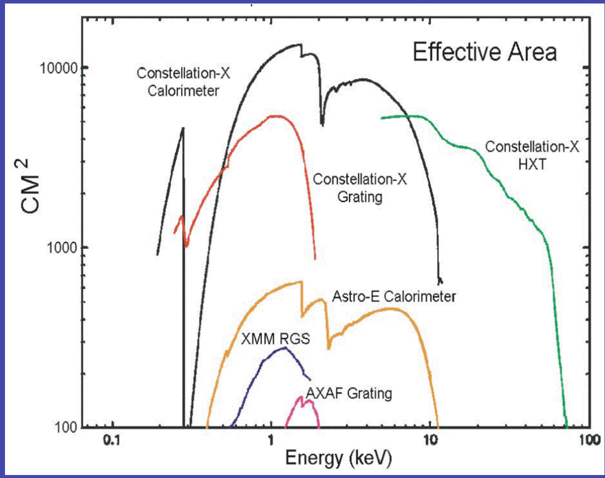
Instrumental Response: Effective Area

R(I,E) is the instrumental response, and includes the Effective Area:

- Effective Area: This gives the amount of collecting area of the mirror+detector system; this usually depends on the position of the source relative to the "optical axis" of the imaging system and the energy of the emission



High Resolution Mirror Assembly on Chandra

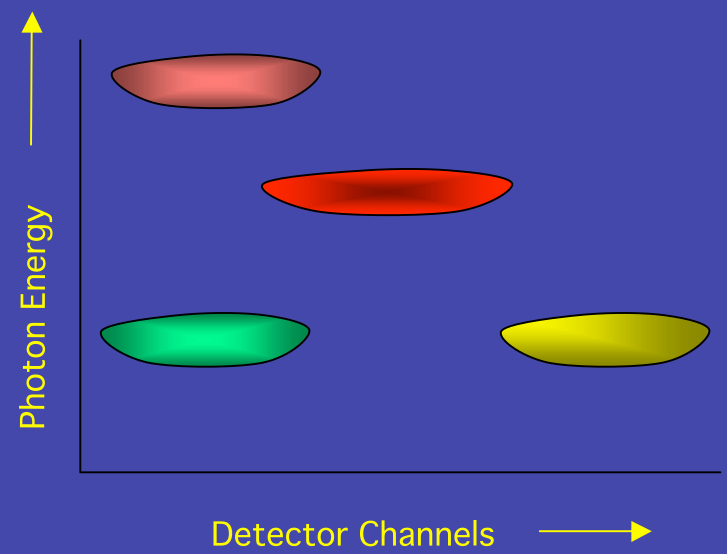


Constellation-X effective area

Instrumental Response: Photon Redistribution

R(I,E) also includes the Photon Redistribution:

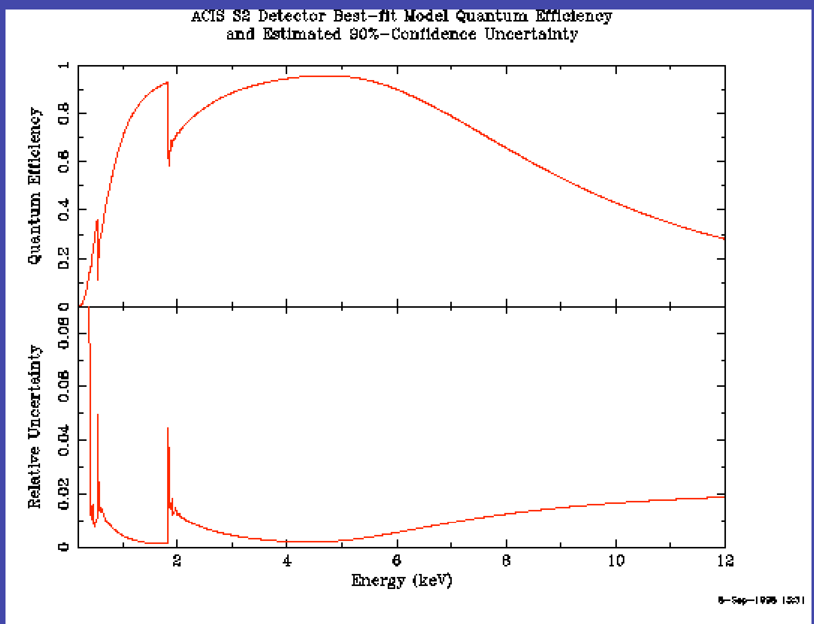
• Photon Redistribution: This quantity gives the probability that a photon of energy *E* will be detected in a detector channel *I*. This is usually given as a matrix of photon probabilities



Instrumental Response: Quantum Efficiency

Detectors don’t detect every event incident on them

Quantum Efficiency: the ratio of the number of incident photons on the detector to the output number of events



QE of Chandra/ACIS

Background

Background are events produced in a detector which are not associated with the astrophysical source of interest.

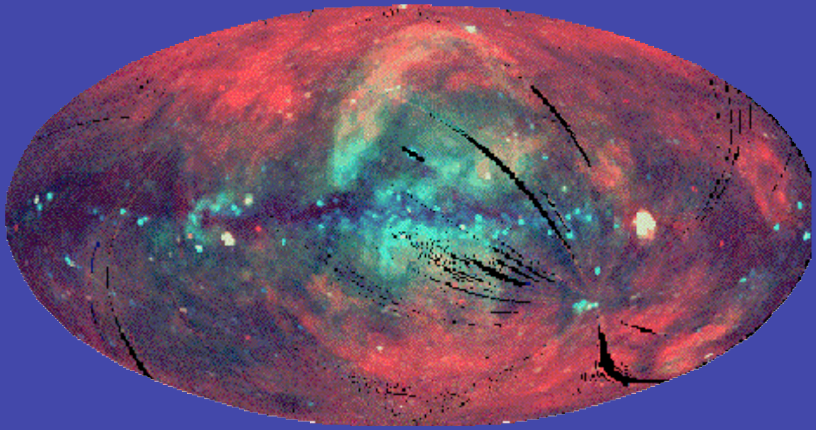
Background can be severe: for example for each gamma-ray photon detected by GLAST, GLAST detects about 80 background events.

$$C(I) = \int_{t_2}^{t_1} \int_0^{\infty} f(E) dt R(I,E) dE + \boxed{B(I)}$$

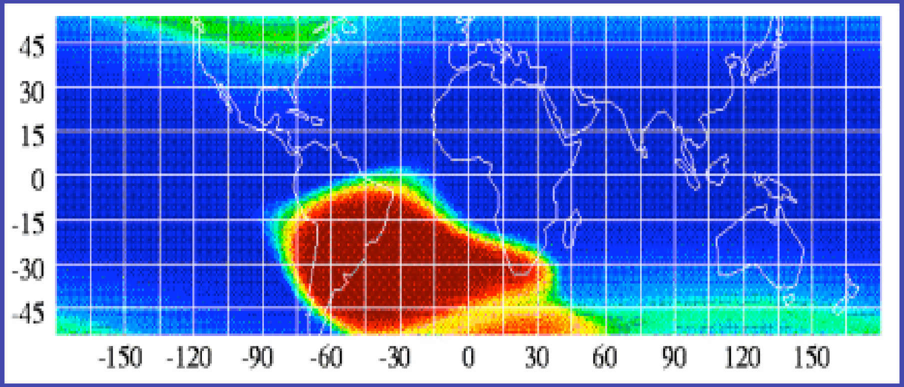
Sources of Background

Background is in general produced by

- charge particles (protons and electrons) in near-earth space environment (esp. in South Atlantic Anomaly, or SAA)
- EM radiation
- Activity within the instrument (CCD dark current)
- Cosmic rays
- other cosmic sources



Cosmic X-ray background: ROSAT All-Sky Survey



Particle Background: SAA

The Background Limit

The amount of background places a limit on the brightness of sources which can be detected. A source needs to be significantly brighter than the background before it can be identified.

This limiting brightness is called the background limit.

Background Elimination

Background can be removed in a number of ways:

- shielding - stops background from reaching the detector
- anti-coincidence rejection - background event is seen in the detector, but an additional (simple) detector helps identify the event as background
- subtraction - the background event is passed to the user in the analysis stage, and the user tries to correct for background in a statistical way - for example, by approximating the background in an observation by the (scaled) amount of background in a neighboring observation.

Data Products

Once events are cleaned, the event data are turned into data products for one or more analysis tasks. Data products include:

- photon event files - a table of all detected events, along with observed parameters (time of arrival, position, energy)
- images - a 2-d matrix giving the number of events in a given pixel. Can be in spatial coordinates (position in some coordinate system) or temporal coordinate system.
- spectra - a table giving the histogram of event energies for a given region of the detector

Data Analysis

Typical analyses include:

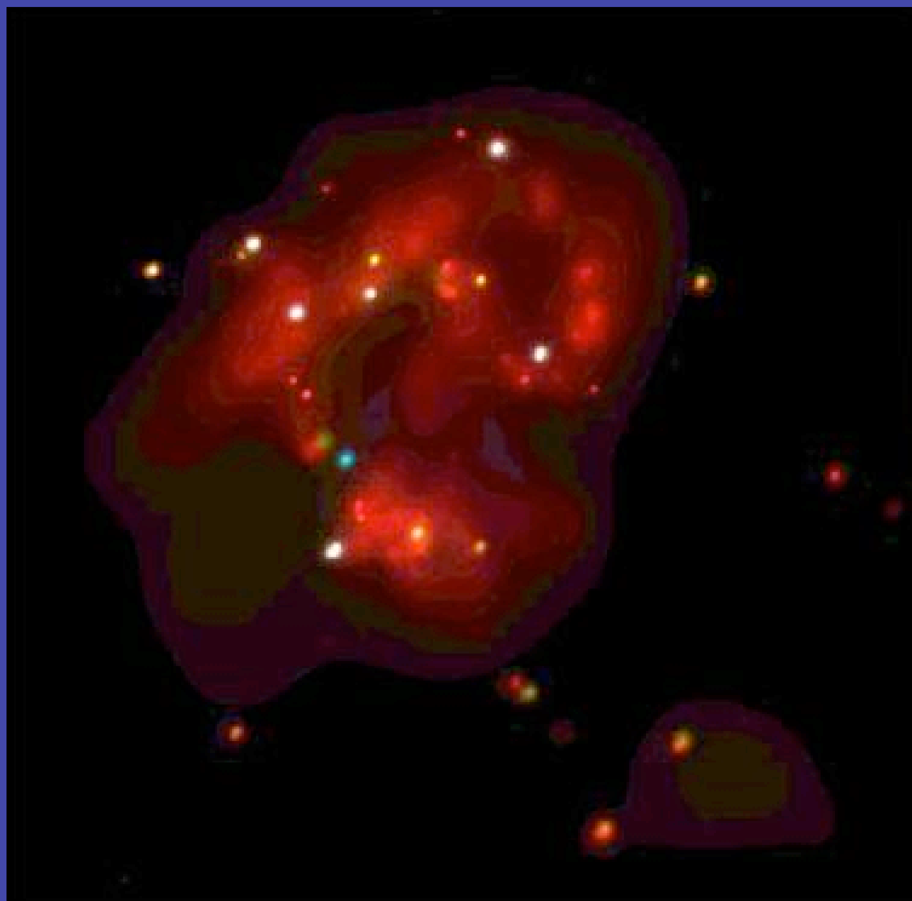
- spatial analyses: determine source locations, brightnesses, counterparts, organization, and amount of confusion
- timing analyses: look for periodic signals
- spectral analyses: determine source of emission, physical properties

Spatial Analysis

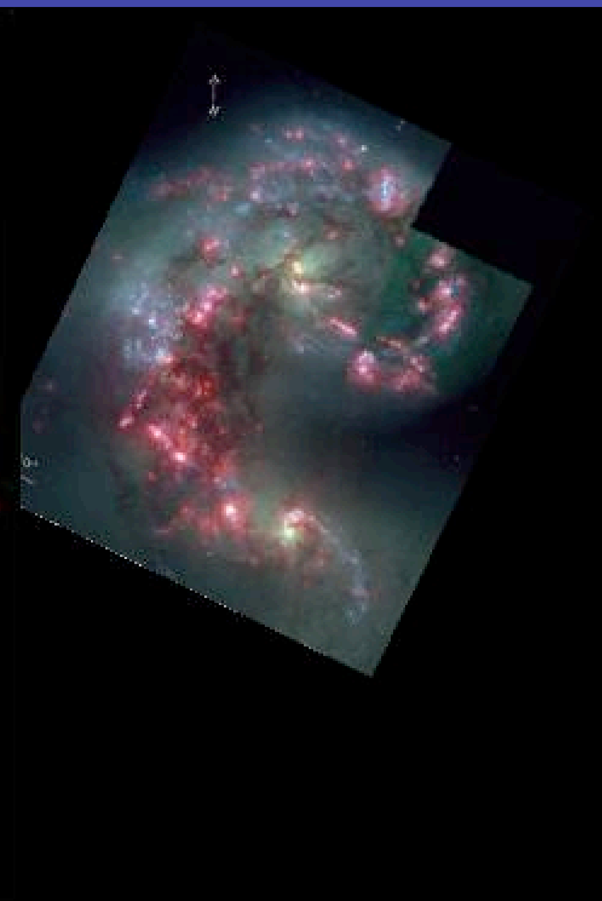
Images help address these questions:

- What is the Luminosity Function (number of sources in a given luminosity band) for a given (physically connected) region of the sky?
- How does the emission at different photon energies compare?
- What are the identities of the sources?
- How are different source categories organized spatially?
- One source or two (or more)? Resolvable?

Sample Images



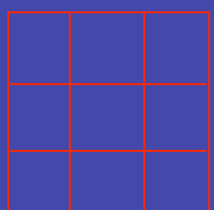
Chandra X-ray Image of the Antenna Galaxy



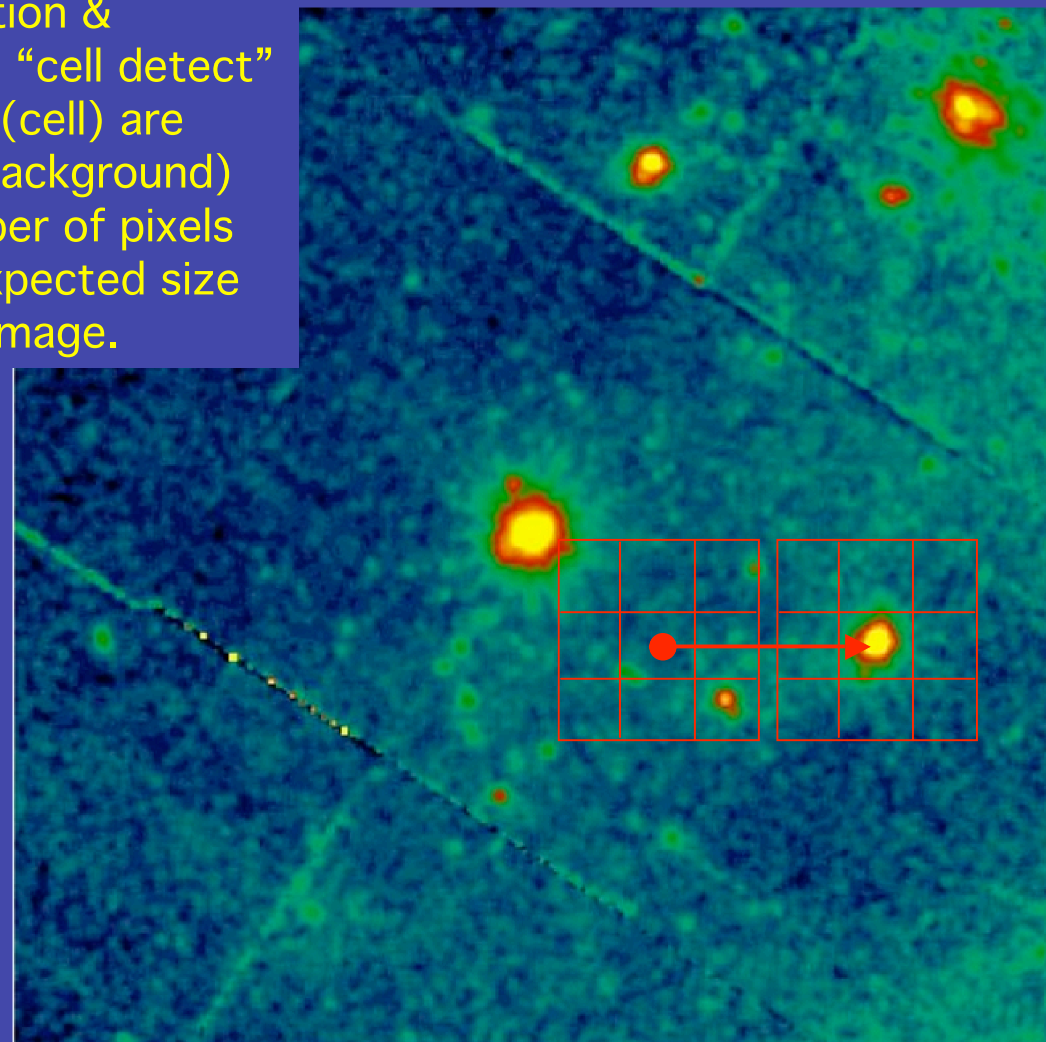
HST optical image of the Antenna Galaxy

Source Detection

Want to determine source position & brightness. Simplest method is “cell detect” in which counts in an inner box (cell) are compared to counts in outer (background) boxes. Cells are a certain number of pixels in width, width can vary with expected size of the psf. Cells “slide” across image.



the cell



Cell Detect Algorithm

S	=	Expected total source counts,
B	=	Expected background counts in detect cell,
C	=	Total counts in square detect cell of size d
T	=	Total counts in detect cell plus surrounding background frame (a b x b pixel box),
Q	=	T - C,
α	=	Integral of point spread function over detect cell,
β	=	Integral of point spread function over detect cell plus surrounding background frame,

then

$$C = \alpha S + B$$

$$T = \beta S + \left(\frac{b}{d}\right)^2 B.$$

$$B = \frac{(\alpha - \beta)C + \alpha Q}{\alpha b^2 d^{-2} - \beta}$$

$$\sigma_B^2 = \frac{(\alpha - \beta)^2 \sigma_C^2 + \alpha \sigma_Q^2}{(\alpha b^2 d^{-2} - \beta)^2}$$

$$S = \frac{C(b^2 - d^2)d^{-2} - Q}{\alpha b^2 d^{-2} - \beta}$$

$$\sigma_S^2 = \frac{\sigma_C^2 (b^2 - d^2)^2 d^{-4} + \sigma_Q^2}{(\alpha b^2 d^{-2} - \beta)^2}$$

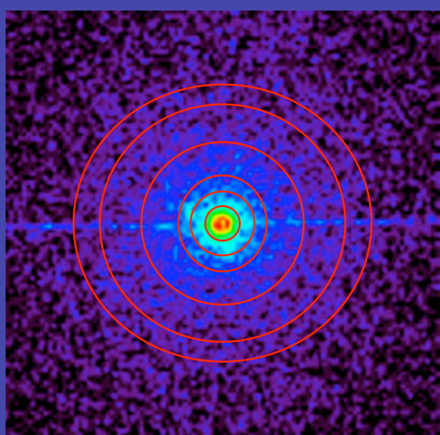
$$\boxed{SNR = \frac{S}{\sigma_S} = \frac{C(b^2 - d^2)d^{-2} - Q}{\sqrt{\sigma_C^2 (b^2 - d^2)^2 d^{-4} + \sigma_Q^2}}}$$

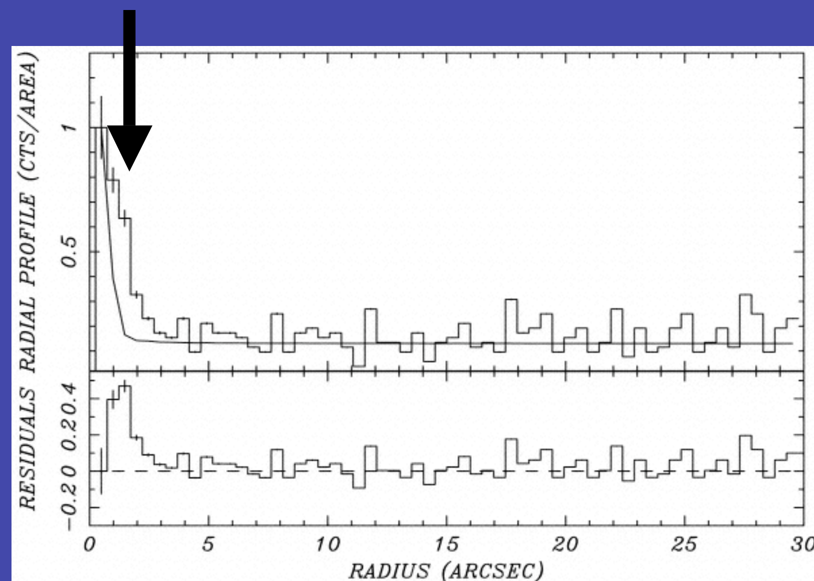
see http://cxc.harvard.edu/ciao/download/doc/detect_html_manual/intro_compare.html#chap:cell_theory

Source Extent

Is a source pointlike or composed of different emission regions? Extract counts in annuli centered on the source, and compare to the observed point spread function:

Excess counts suggests presence of an X-ray emitting nebula around this pulsar





Concerns: normalization (pileup), centering

Timing Analysis

Events are placed into time bins in order to generate the variation of source count rate (or brightness) vs. time.

bin size needs to be large enough for sufficient number of counts per bin but small enough so that real variation is not hidden.

The "lightcurve" (brightness vs. time) can be analyzed to determine whether a source varies, and if so, is the variation periodic. If periodic, can derive physical parameters (size, mass).

$$X_i = X(t_i) = X_s(t_i) + B(t_i)$$

Epoch Folding

The simplest period determination is via epoch folding.

for a variable phenomena, choose a test period and fold the data by calculating a phase as

phase=ϕ=(time-E)/P &

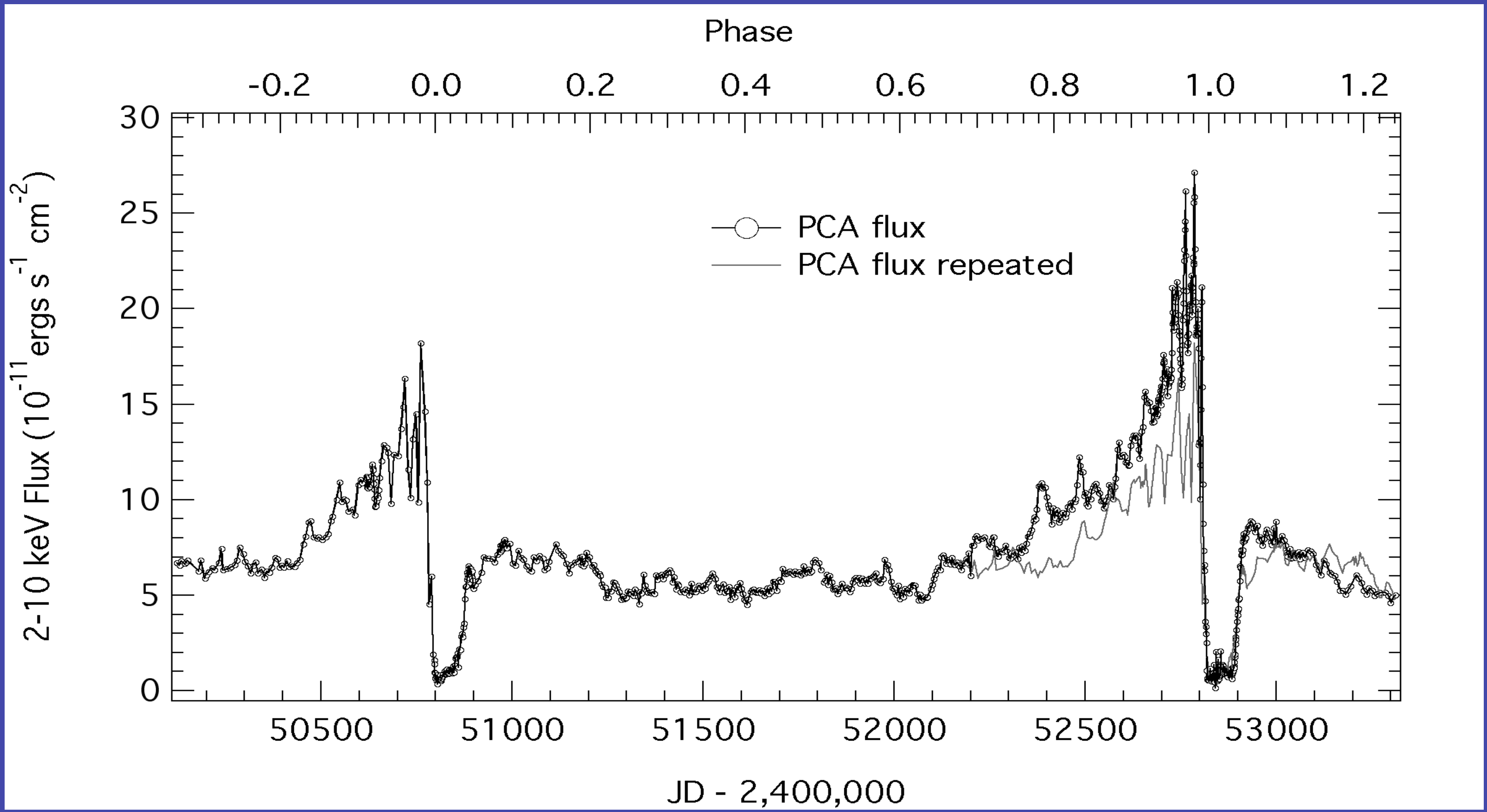
where E is the "epoch" and P the test period, and phases are usually constrained to be between 0 and 1:

$\phi=\phi$-(integer part of ϕ)

$X(t_i) \Rightarrow X(\phi_i)$

Vary period until $X(\phi_i)$ at given phases match.

Sample Epoch Fold Analysis:



X-ray emission from Eta Car measured by RXTE vs time & ϕ

Fourier Techniques

Variable sources can be expressed as some combination of sines and cosines

A Periodogram (a Fourier transform for discretely sampled data) is defined as

$$P_x(\omega) = \frac{1}{N_0}\left[\left(\sum_j X_j \cos \omega t_j\right)^2 + \left(\sum_j X_j \sin \omega t_j\right)^2\right]$$

Power at test frequency ω

data at given times t_j

number of data points

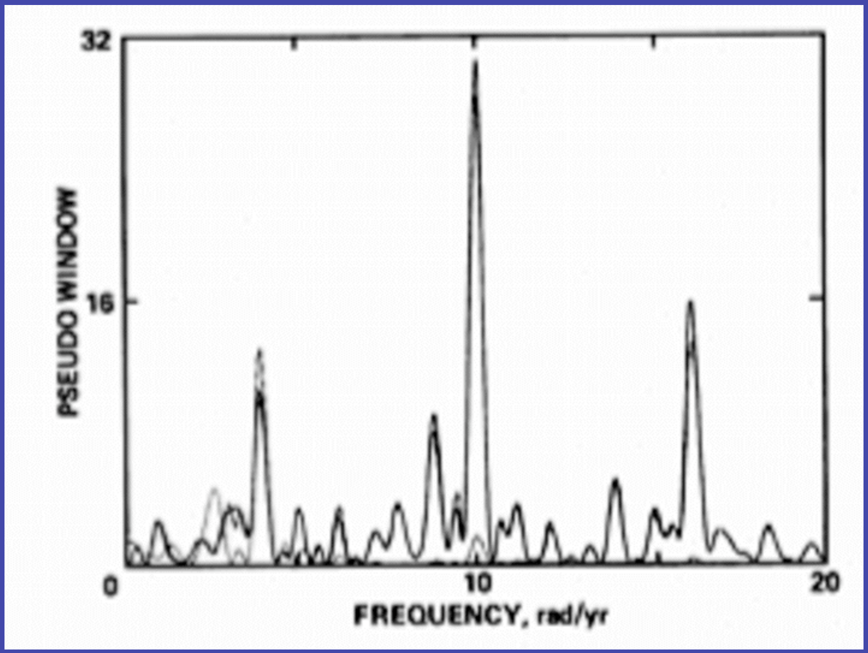
Scargle, 1982, ApJ, 263, 835

Fourier Techniques & Sampling

Plot Px(ω) vs ω

If the phenomenon under study varies sinusoidally with a period P_0, $P_x(\omega)$ will have a peak at the frequency $\omega_0=2\pi/P_0$

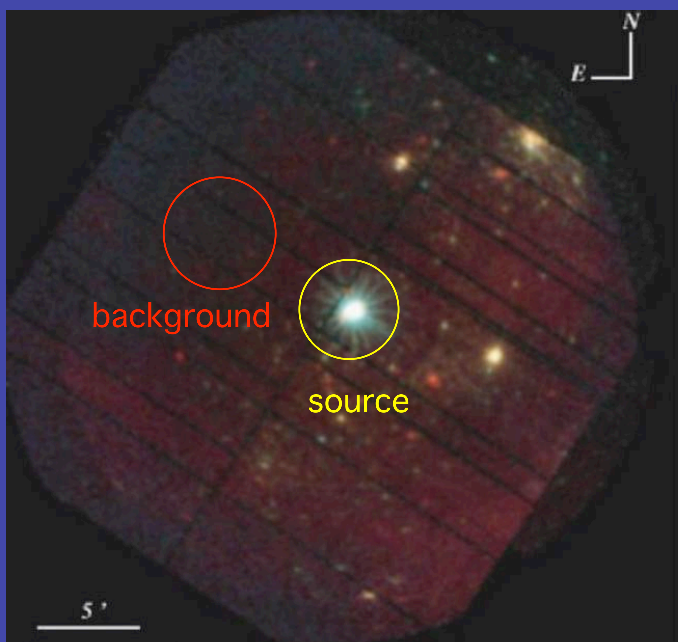
This works best for evenly sampled data (for example a few ksec pointing at a pulsar with a pulse period of 1 sec)
For data with gaps in the sampling (long period binary, for example), need to use a Discrete Fourier Transform.



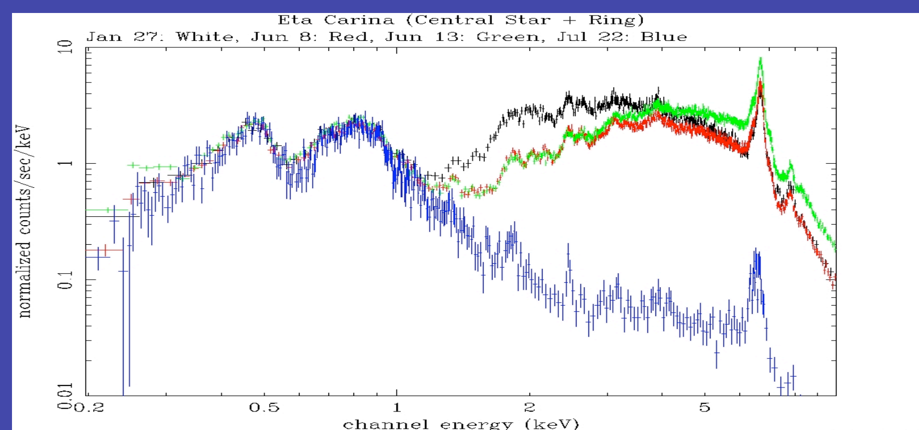
Spectral Analysis

in x-ray & gamma-ray astronomy, spectral analysis involves the extraction of a source (+background) counts from a list of photon events, and the comparison of a model spectrum (modified by the detector response) to those data.

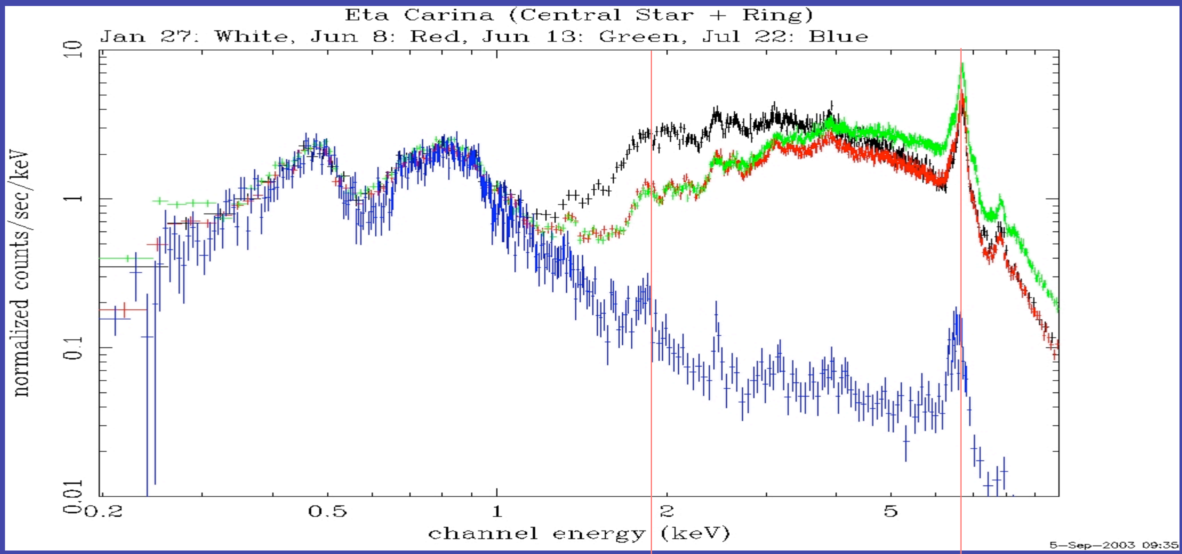
Typically spectral resolution is low ($E/\Delta E<1000$) so there can be much uncertainty in determining appropriate model parameters.

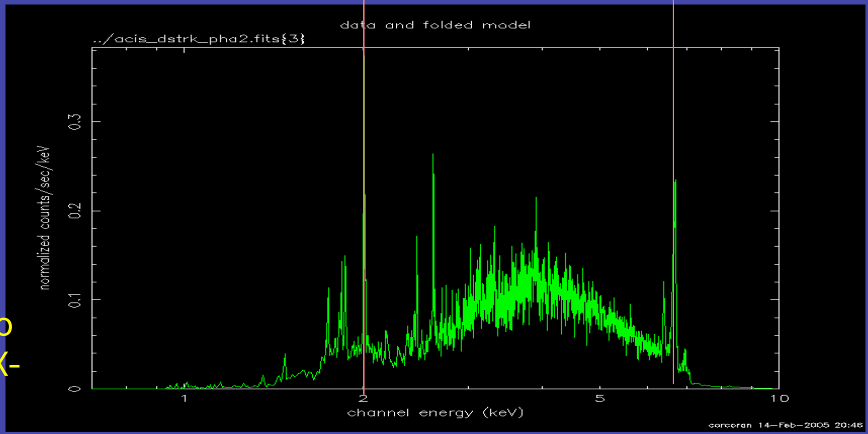


An X-ray spectral image where color represents photon energy: red - low energy; blue - high energy



CCD vs. Gratings





Higher spectral resolution of grating data provideaccess to important diagnostics using X-ray spectral lines.

Modeling a Spectrum

use estimate of instrumental response and background along with a physical emission model to compare to observed source counts

$$C(I) = \int_{t_2}^{t_1} \int_0^{\infty} f(E) dt R(I,E) dE + B(I)$$

$$C(I) - B(I) = \int \int f(E,t) dt R(I,E) dE$$

$$C(I) = \int \int (f(E,t) - B(E,t)) dt R(I,E) dE$$

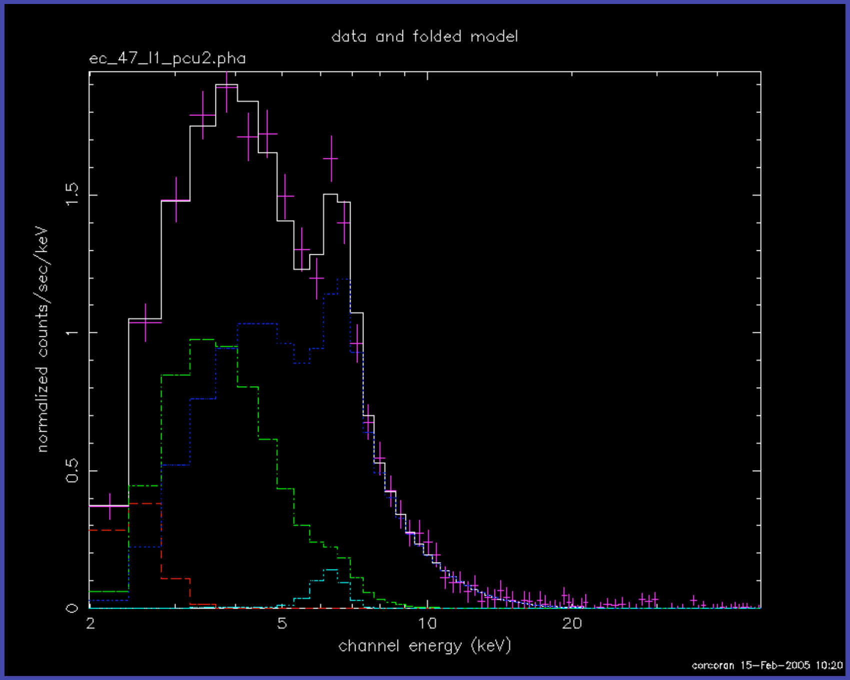
In the comparison of the "convolved" model to the data goodness-of-fit is defined by simple statistical tests

Commonly Used Models

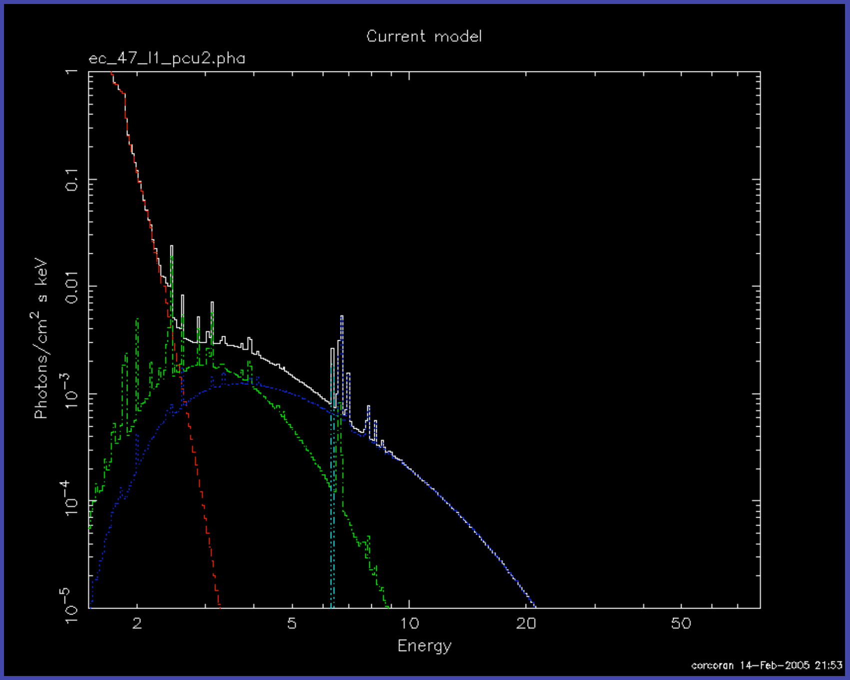
common types of spectral models, used in different physical situations:

- emission from shocked plasma: dominated by thermal bremsstrahlung + line emission from collisionally-ionized recombining atoms (Raymond-Smith, Mewe-Kaastra-Liedahl)
- non-thermal models: synchrotron emission, usually described by a power law in photon energy (or photon number)
- blackbody emission: hot objects like neutron stars and black holes
- other more specific models: non-equilibrium shocks, specific accretion disk models, cooling flows of gas from clusters of galaxies

Testing models



Data and model components



Emission models: 3 plasmas in collisional-ionization equilibrium plus absorption plus a narrow emission line

Goodness of Fit

Adequacy of model is judged statistically using χ^2 statistic or some variant

$$\chi^2 = \sum (C(I) - C_p(I))^2 / \sigma(I)^2 ; \sigma(I) = \sqrt{C(I)}$$

chi-squared statistic

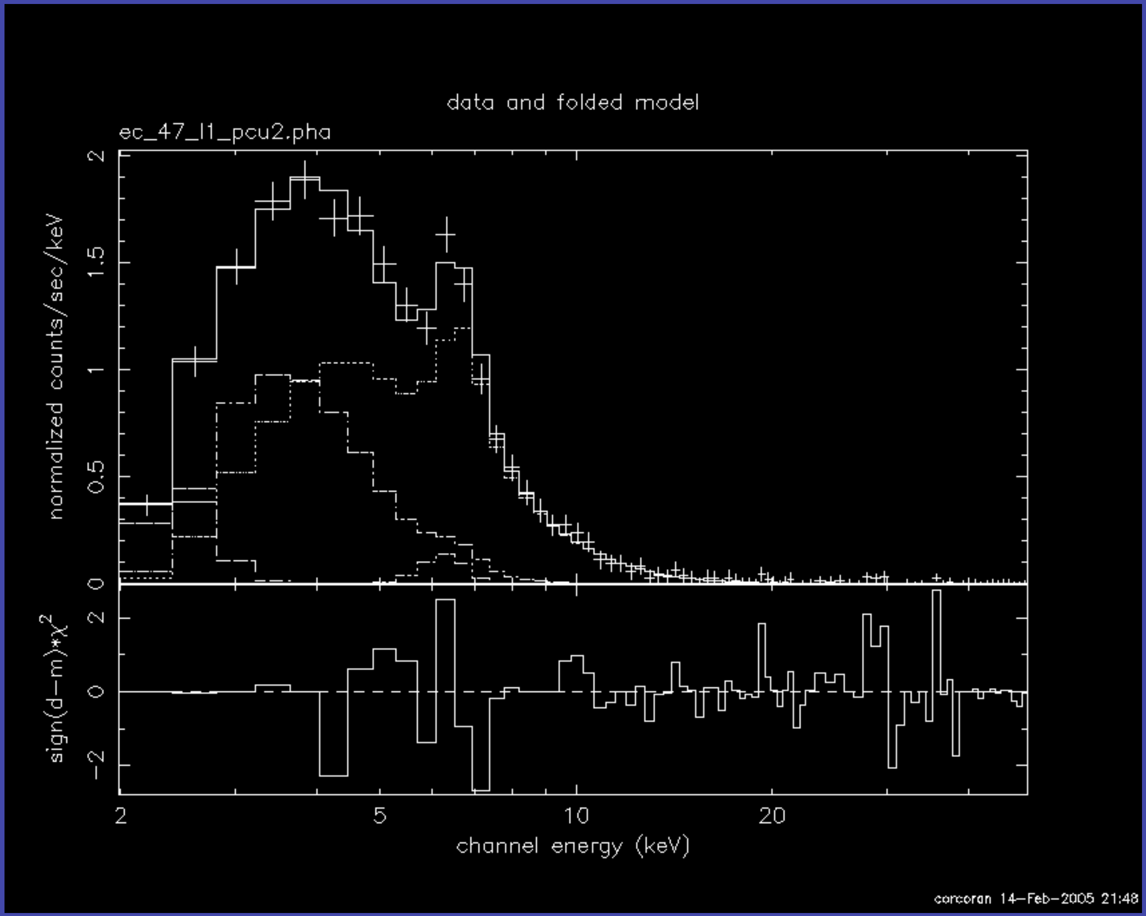
observed counts in bin

predicted counts in the bin

uncertainty in the bin

Poissonian statistics

Goodness of Fit



ν = number of degrees of freedom (number of channels - number of model parameters)

want $\chi^2/\nu = 1$

for a statistically acceptable fit

Parameter Estimation

Once a statistically acceptable fit is achieved the uncertainty of the parameters needs to be estimated

This is usually done by finding a confidence interval, in which all other parameters are held fix and the parameter of interest is varied until the χ^2 value changes by an amount which yields the desired level of confidence.

For example a fit to a spectrum using a blackbody model plus ISM absorption has 3 parameters: temperature+brightness of the BB, and column density of absorption.

Confidence	Parameters		
	1	2	3
0.68	1.00	2.30	3.50
0.90	2.71	4.61	6.25
0.99	6.63	9.21	11.30

$\Delta\chi^2$ table

Conclusions

before conclusions can be drawn from astrophysical observations:

- effects of instruments on incident radiation must be understood
- Data often need correction for detector biases, imperfections and non-linearities
- Compare model to data not vice-versa
- sources of background must be understood and removed as much as possible

Statistics ≠ physics